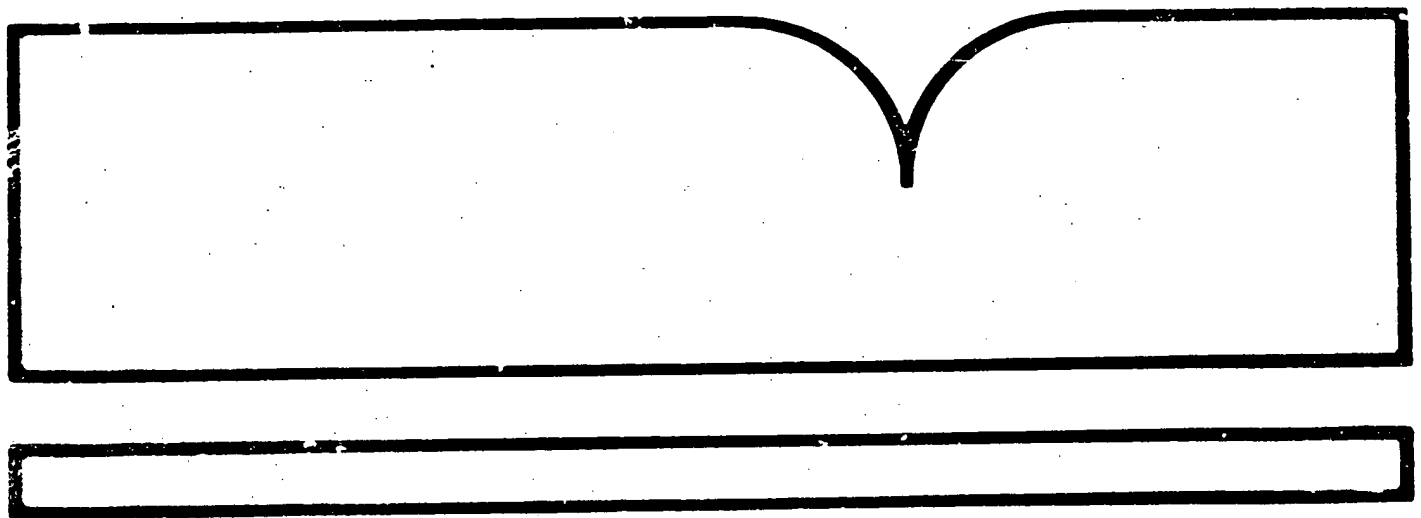PB89-186233

Nucleotide and Amino Acid Sequences of the
Four Variable Domains of the Major Outer
Membrane Proteins of 'Chlamydia trachomatis'

(U.S.) Department of Health and Human Services
Washington, DC

17 Mar 89

Abstract: Based on the immunological and biological relationships among Major Outer
Membrane Proteins Variable Domains (MOMP VD) structure and function, and the fact
that the MOMP VDs appear to be the major variable sequences between serovars, the
present inventors identified the nucleotide and amino acid sequences of the MOMP
VDs I-IV of C. trachomatis serovars Ba, D, E, F, G, H, I, J, K, and L3, and
confirmed those of serovars B, A, C, L1 and L2. The present invention will allow
the production of reagents and methodologies applicable in the development of new
diagnostic tests for C. trachomatis infections and serological tests for
serotyping. Specifically, species or sero-type specific oligopeptides conjugated to
an array of reporter or detector groups could be used in various conventional non-
immunologic assays including standard hybridization techniques.

APPLICATION FOR UNITED STATES PATENT

HHS

Inventor(s):
Harlan D. CALDWELL
Yuan YING
You-Xun ZHANG
Nancy G. WATKINS

Invention:
NUCLEOTIDE AND AMINO ACID SEQUENCES OF THE FOUR VARIABLE DOMAINS OF THE MAJOR OUTER MEMBRANE PROTEINS OF CHLAMYDIA TRACHOMATIS

Cushman, Darby & Cushman
1615 L Street, N.W.
Eleventh Floor
Washington, D.C. 20036-5601
Attorneys
Telephone: 861-3000

SPECIFICATION

CDC-100   5/88

BACKGROUND OF THE INVENTION

In the United States, urogenital infections with *Chlamydia trachomatis* are the leading sexually transmitted disease with an estimated 10 million new cases reported each year. The urogenital infections result in involuntary infertility in 100-200,000 women each year. It is estimated that equal or greater numbers of these infections occur in European countries. In addition, ocular infection with *C. trachomatis* results in blinding trachoma which afflicts approximately 500 million individuals from developing countries. Improved diagnostic methodologies and vaccines or immunoprophylactic preventative measures are need to manage and control the disease.

Chlamydia trachomatis isolates occur as 15 distinct serovars. Based on serological relatedness, these 15 serovars have been divided into three serogroups; B serogroup (serovars B, Ba, D, E, L1 and L2), intermediate serogroup (serovars F, G, K and L3), and C serogroup (serovars A, C, H, I and J). The antigen that confers serovar and serogroup-specificities to chlamydia is the major outer membrane protein (MOMP), and protective immunity developed during chlamydial ocular infection is thought to be directed at serovar and serogroup MOMP determinants.

The genes encoding the MOMP of the *C. trachomatis* serovars A, B, C, L1 and L2 have been cloned and sequenced (Pickett et al., FEMS Microbiol. Lett. 42:185-190 (1987); Stephen et al., J. Bacteriol. 168:1277-1282 (1986); and Stephens et al., J. Bacteriol 169:3879-3885 (1987); Baehr et al., Proc. Natl. Acad. Sci. USA, 85:4000-4004

(1988)). Comparative analysis of their amino acid
sequences show the MOMP genes to be highly conserved
structures that contain four evenly spaced domains
whose sequences vary among the different serovars.

5 The locations of these nucleotide and amino acid
sequence variable domains (VDs) are:

VDI   - nucleotides 256-315 and residues 64-83;
VDII   - nucleotides 481-546 and residues 139-160;
VDIII - nucleotides 736-777 and residues 224-237;
10 VDIV   - nucleotides 928-1017 and residues 288-317.

Epitope mapping has shown that three of the
four VD domains (I, II and IV) contain contiguous
antigenic determinants which elicit the formation of
serovar, subspecies (determinants common to three or

15 more serovars within a serogroup), serogroup or
species-specific antibodies.

Variable domains I and II, which demonstrate
the greatest amount of inter serogroup sequence
variation, are the locations of the serovar-specific

20 determinants. Variable domain IV is the largest of
the MOMP VDs and is located near the C-terminus of
the protein. VDIV is the location of subspecies and
serogroup antigenic determinants, as well as a
highly conserved species-specific antigenic

25 determinant. Antigenic determinants have not been
mapped to VDIII, the smallest and least variable
domain of the MOMP genes.

Variable domains I, II and IV of the MOMP of _C. trachomatis_ protrude from the cell surface towards the external environment as demonstrated by their susceptibility to cleavage by trypsin, and

5   their accessibility to antibody binding.   Trypsin cleavage in both VDII and VDIV, but not within VDIV alone, decreases chlamydial attachment to HeLa cells suggesting that these domains, or conformational MOMP structures that are dependent on the integrity

10 of these domains, may function as a chlamydial ligand.

Based on the immunological and biological relationships among MOMP VD structure and function, and the fact that the MOMP VDs appear to be the

15 major variable sequences between serovars, the present inventors sought to identify the nucleotide and amino acid sequences of the MOMP VDs I-IV of _C. trachomatis_ serovars Ba, D, E, F, G, H, I, J, K and L3, and to confirm those of serovars B, A, C, L1 and

20 L2.

SUMMARY OF THE INVENTION

The amino acid sequences of the major outer membrane protein genes from _Chlamydia trachomatis_ serovars A, B, C, L1 and L2 have been shown to be

25 conserved with the exception of four variable

domains in which major neutralizing and serotyping antigenic determinants have been located. Using oligonucleotide primer extension sequencing of MOMP mRNA, the nucleotide sequences of the four MOMP VDs of the remaining 10 C. trachomatis serovars were determined by the present inventors and the amino acid sequences deduced therefrom. The sequencing technique was rapid and required minimal amounts of total RNA, i.e., only 35 μg of RNA from chlamydial infected HeLa cells was needed to sequence all 4 variable domains for a given serovar. Use of the disclosed technique also circumvented the more standard and laborious approach of molecular cloning and direct DNA sequencing which had been employed in the past to determine the sequences of MOMP gene to serovars A, B, C, L1 and L2.

As noted above, the nucleotide sequences were employed to deduce the corresponding amino acid sequences. Comparative analysis of the amino acid sequence homology of the four variable domains permitted the present inventors to separate the fifteen serovars into three serogroups:

Group 1 - serovars B, Ba, D, E, L1 and L2;
Group 2 - serovars G and F;
Group 3 - serovars A, C, H, I, J, K and L3.

The four variable domains were also analyzed for immunogenicity based on the hydrophilicity and

charge values of each domain (Baehr et al., supra). The MOMP VDs with the greatest total hydrophilicity and charge values were found to be the location of antigenic determinants recognized by MOMP specific monoclonal antibodies (Stephens et al., supra; Baehr et al., supra).

The nucleotide, amino acid sequences and hydrophilicity/charge value analyses are advantageous because they will assist in the selection of appropriate MOMP antigenic determinants to be used in the construction of synthetic peptides, subunits or recombinant chlamydial vaccines. Examples of such recombinant vaccines include infectious enteric vectors (avirulent Salmonella typhimurium and/or Sabin Type I polio virus) expressing MOMP variable domain sequences as antigenic chimeras or hybrid molecules.

The present invention will allow the production of reagents and methodologies applicable in the development of new diagnostic tests for C. trachomatis infections and serological tests for serotyping. Specifically, species or sero-type specific oligopeptides conjugated to an array of reporter or detector groups could be used in various conventional non-immunologic assays including standard hybridization techniques.

Likewise, reagents for use in immunologic or

serologic tests could be prepared based on the

sequences presently disclosed; the conserved VD-IV

peptide sequence TTLNTTIAG would permit species

specific identification, while the non-conserved

sequences would enable one to make sero-type

specific identification. It is contemplated by the

present inventors that such reagents would be

suitable for use in standard immunologic procedures

usch as ELISA and radioimmunoassay techniques.

The invention will be more fully described in

the detailed description and drawings which follow.

DESCRIPTION OF THE DRAWINGS

Fig. 1. Location of the MOMP gene sequences

to which synthetic complementary oligonucleotides

were constructed and used for primer extension mRNA

sequencing of MOMP VDs. The MOMP gene sequences

identified by single lined boxes are those to which

synthetic oligonucleotides were constructed. The

double lined boxes identify the four variable domain

regions of B, L2, A, C, and F MOMP genes.

Fig. 2. Sequencing of serovar B and A MOMP

VDII by primer extension of MOMP mRNA. Total RNA

was isolated from HeLa 229 cells infected with

serovars B and A and annealed to kinased

oligonucleotides BFP-2 and CP-2 (Table 1),
respectively. The oligonucleotides were
complementary to the non-coding strand of MOMP gene
sequences 50 nucleotides downstream 3' of VDII. The

5   RNA templates were transcribed with reverse
transcriptase, and the sequences of the transcribed
DNAs determined by the dideoxynucleotide chain
termination method. Audioradiographs were read to
depict + strand MOMP DNA sequences. The nucleotide

10   numbers shown for each MOMP are the beginning and
end of the sequence encoding VDII of each protein.
The results shown are representative of sequencing
results obtained for all other serovars' VDs.

        Fig. 3. (A, B, C and D) The nucleotide

15   sequences of MOMP VDs I, II, III and IV,
respectively, of the 15 C. trachomatis serovars.
The boxed regions of the sequences identify each of
the four VDs. Serovars B, F and C MOMP sequences
were used as prototypes for comparative sequence

20   analysis. The nucleotide positions shown are for
serovar B MOMP. Exact nucleotide positions for
other MOMP VD sequences differ slightly because of
insertions and/or deletions within the VDs.

        Fig. 4. (A, B, C and D) The deduced amino

25   acid sequences of the MOMP VDs 1, 2, 3 and 4,
respectively, of the 15 C. trachomatis serovars.

The boxed regions of the sequences identity each of the four VDs. Serovars B, F and C MOMP amino acid sequences were used as prototypes for comparative sequence analysis.

5   Fig. 5. C. trachomatis serovar B MOMP variable domain sequence.

Fig. 6. C. trachomatis serovar Ba MOMP variable domain sequence.

Fig. 7. C. trachomatis serovar D MOMP

10   variable domain sequence.

Fig. 8. C. trachomatis serovar E MOMP variable domain sequence.

Fig. 9. C. trachomatis serovar L1 MOMP variable domain sequence.

15   Fig. 10. C. trachomatis serovar L2 MOMP variable domain sequence.

Fig. 11. C. trachomatis serovar F MOMP variable domain sequence.

Fig. 12. C. trachomatis serovar G MOMP

20   variable domain sequence.

Fig. 13. C. trachomatis serovar C MOMP variable domain sequence.

Fig. 14. C. trachomatis serovar A MOMP variable domain sequence.

25   Fig. 15. C. trachomatis serovar H MOMP variable domain sequence.

Fig. 16. C. trachomatis serovar I MCMP variable domain sequence.

Fig. 17. C. trachomatis serovar J MOMP variable domain sequence.

5   Fig. 18. C. trachomatis serovar K MOMP variable domain sequence.

Fig. 19. C. trachomatis serovar L3 MOMP variable domain sequence.

DETAILED DESCRIPTION OF THE INVENTION

10   All references cited below are specifically incorporated into the specification by reference.

I.   MATERIALS AND METHODS

A. Chlamydiae.

C. trachomatis serovars: A/Har-13, B/TW-
15   5/OT, Ba/AP-2, C/TW-3/OT, D/UW-3/Cx, E/Bour, F/IC-
Cal-13, G/UW-57/Cx, H/UW-4/Cx, I/UW-12/Ur, J/UW-
36/Cx, K/UW-31/Cx, LGV/L1-440, LGV/L2-434 and
LGV/L3-404 were employed herein and were immunotyped
by the micro-immunofluorescence (micro-IF) procedure
20   of Wang et al. (J. Infect. Dis. 152:791-800 (1985)).
Serovars A, E and F were obtained from the American
Type Culture Collection (ATCC, Rockville, MD). All
other serovars were obtained from Dr. Cho-Chou Kuo,

Department of Pathobiology, University of
Washington, Seattle, WA.

    B.   <u>RNA extraction</u>.

    Approximately 2×10^8 HeLa 229 cells grown in
5   stationary culture were each infected with one
chlamydial serovar. Infected HeLa cells were
removed from the culture flasks with trypsin 24-
hours post-infection for serovars L1, L2 and L3, and
36-hours post-infection for all other serovars.
10   Cell suspensions were then pooled, centrifuged and
washed twice in cold (4°C) phosphate-buffered saline
(PBS), 0.02 M sodium phosphate, 0.15 M NaCl, pH 7.2.
Cells were lysed in 5 ml of cold guanidine
isothiocyanate lysis buffer (Current Protocols in
15   Molecular Biology, Wiley Inter Science, p. 4.2.4
(1987)). The resulting suspension was then gently
refluxed through an 18 gauge needle 15-20 times to
shear the DNA.

    Total RNA was extracted (2x) from the
20   solution using hot acidic phenol (65°C, pH 5.0),
followed by two extractions with chloroform/isoamyl
alcohol (24:1 vol/vol). The RNA from each pool was
precipitated with two volumes of cold ethanol (-20°
C). The precipitate was evaporated to dryness using
25   a Speed-vac (Bachofer, Reutlingen, FGR). The RNAs

were resuspended in glass distilled water at a

concentration of 20-40 μg/ml and stored at -70°C.

C. Synthetic oligonucleotides.

Oligonucleotides (21 mers) complementary to

5   the + strand MOMP DNAs flanking the 3' end of each

VD were synthesized using a SAM1 automated synthesis

instrument (Biosearch, San Rafael, CA) by the N-

methylimidazole phosphotriester method described by

Grayston et al. (J. Infect. Dis. 132:87-104 (1975))

10   and according to the manufacturer's instructions.

De-blocked oligonucleotides were evaporated to

dryness using a Speed-Vac, and were then dissolved

in distilled water at a concentration of 10-30 mg/ml

and stored at -20°C. Oligonucleotides were

15   constructed according to the published sequences of

the A, B, C, L2 MOMP gene serovars (Stephens et al.,

J. Bacteriol. 168, supra and Baehr et al., supra),

as well as the unpublished F serovar sequence which

was determined by the present inventors.

20   The MOMP gene sequences to which synthetic

complementary oligonucleotides were constructed, and

their location in the MOMP genes are shown in Fig.

1. The oligonucleotide number, its sequence, and

the MOMP serovar and VD for which the

25   oligonucleotide was used for primer extension mRNA

sequencing are summarized below in Table I.

oligonucleotide was used for primer extension mRNA

sequencing are summarized below in Table I.

TABLE 1. Oligonucleotides Used in Primer-extension

mRNA Sequencing of C. trachomatis MOMP VDs

Oligonucleotide number	Oligonucleotide sequence	Serovars and their VDS sequencing with oligonucleotide primers
BP-1	5'-ATCCCAAATATTCAATGCCAT-3'	VDIs of B,Ba,D,E,G,L1,L2
CP-1	ATCCCAGATATTTAATGCCAT	VDIs of A,C,H,I,J,K,L3
FP-1	ATCCAAATATTCAATGTCAT	VDIs of F,G
BFP-2	TCCACATTCCCACAAAGCTGC	VDIIs of B,Ba,D,E,F,G,L1,L2
CP-2	CCCACATTCCCAGAGAGCTGC	VD1Is of A,C,H,I,J,K,L3
BP-3	TCTGTAAGAGAGAGCTAAACT	VDIIIs of B,Ba,D,E,L1,L2
CP-3	AGTGAACATATTTAATCTGTA	VDIIIs of A,C,H,I,J,K,L3
FP-3	AGTGAACATATTGAGTCTGTA	VDIIIs of F,G
P-4-1	AATACCGCAAGATTTTCTAGA	VDIVs of A,B,Ba,C,D,E,F,H,I, J,K,L1,L2
P-4-2	TCCTACTGCAATACCGCAAGA	VDIVs of L3,G, (ND*)
P-4-3	TTTTCTAGATTTCATCTTGTT	VDIV of J, (ND*)

D. Primer extension and DNA sequencing.

The dideoxynucleotide chain terminating
method of Sanger et al. (Proc. Natl. Acad. Sci. USA
74:5463-5467 (1977) modified for the use with
reverse transcriptase and RNA templates as described
by Lane et al. (Proc. Natl. Acad. Sci. USA 82:6955-
6959 (1985)) was used for sequence determinations.
In summary, oligonucleotides were 5' end labeled
with [^{32}P]dATP (3,000 Ci/mmol - New England Nuclear,
Boston, MA) using T4 polynucleotide kinase (10,000
U/ml - New England Biolabs, Beverly, MA) as
described by Maniatis et al. in A laboratory manual.
pp. 1-545 (Cold Spring Harbor Laboratory, Cold
Spring Harbor, NY (1982). Labeled oligonucleotides
were isolated on Sephadex G-25 columns (Boehringer
Mannheim Biochemicals, Indianapolis, IN). The
oligonucleotide primers were then hybridized to MOMP
mRNA templates in 12.5 µl reaction mixtures
containing 2.5 µl 5x annealing buffer (250 mM Tris-
HCl, 300 mM NaCl, 250 mM DTT, and 5 mM EDTA, pH
8.3), 7.5 µl of total RNA (3-10 µg/µl). and 2.5 µl
of kinased primer (2-2.5 µM).

The mixtures were heated at 90°C for 3
minutes, centrifuged and then allowed to cool slowly
to 45°C over a 30 minute period. 2 µl of the
hybridized mixture was then added to each of four
tubes containing 3µl of reaction mixture (250 mM
Tris-HCl, pH 8.3, 300 mM NaCl, 50 mM DTT, 30 mM
MgCl$_2$, 2 mM dNTPs, 0.4 mM ddNTPs and 0.6 - 0.8 U/µl
avian myeloblastosis virus reverse transcriptase
(20,000 U/ml - Pharmacia Fine Chemicals, Inc.,
Piscataway, N.J.). The mixtures were incubated at
45°C for 40 minutes then the reactions were stopped
by evaporating the tubes to dryness in a Speed-vac.

Following evaporation, the samples were resuspended
in 3 μl of 100% formamide containing 0.8% xylene
cyanol and 0.8% bromophenol blue and heated at 95°C
for 3 minutes. The samples were then promptly
loaded on sequencing gels (8% polyacrylamide) to
achieve separation. Following separation the gels
were fixed with methanol, dried then
autoradiographed.

II.   Sequence Analysis
       A. Primer extension and DNA sequencing.
       Using the oligonucleotide primers described
above, the VD-encoding portions of the MOMPs mRNA
were sequenced. A representative sequencing gel is
shown in Fig. 2 wherein MOMP B-VDII and A-VDII are
shown. Oligonucleotides BFP-2 and CP-2 were
employed in this sequencing as primers. The mRNA
derived MOMP sequences depicted are identical to the
previously published genomic DNA sequences for the
MOMP B-VDII and A-VDII (Stephens et al., supra and
Baehr et al., supra), thereby demonstrating both the
feasibility and accuracy of this sequencing method.

       B. Sequence analysis.
       The synthetic oligonucleotides used for
sequencing the MOMP VDs of the 15 C. trachomatis
serovars are those depicted previously in Table 1.
The location of the complementary sequences in the
MOMP genes of serovars B, L2, A, C and F are shown
in Fig. 3. All oligonucleotides were 21 mers and
were complementary to conserved nucleotide sequences
40 to 60 nucleotides downstream (3') from the VD to
be sequenced. Because inter-serogroup (but not
intra-serogroup) nucleotide variation occurs within
these regions, oligonucleotides complementary to

serogroup specific MOMP sequences were constructed for use in the sequencing of the variable domains of the serovars within serogroup. The 40 to 60 nucleotide distance between the VD and the downstream oligonucleotide primer site was necessary to produce high resolution sequences of the VDs.

The nucleotide and deduced amino acid sequences of the 4 variable domains for each of the 15 C. trachomatis serovars are individually depicted in Figs. 5-19. These sequences are also presented comparatively in Fig. 3 (nucleotide sequences) and Fig. 4 (deduced amino acid sequences). The B, F and C serovars' MOMP VDs are used as prototype sequences for comparative analyses of serovar VDs within the B, intermediate and C serogroup, respectively. A discussion of the results of the comparative analyses of the nucleotide and amino acid sequences of the four VDs for each C. trachomatis serovar is presented below.

1. Variable Domain I. VDI contains 60 nucleotides encoding 20 amino acids for all B serogroup serovars, 66 nucleotides encoding 22 amino acids for serovars K, L3 and all C serogroup serovars, and 63 nucleotides encoding 21 amino acids for serovars F and G. The nucleotide sequences of VDIs from B serogroup serovars are highly conserved with only 3 to 5 substitutions per serovar as compared to the B serovar prototype sequence. The deduced amino acid sequences of VDIs of B serogroup serovars are also conserved with only 2 to 4 substitutions per serovar, with the majority of substitutions occurring within a 12 amino acid region (residues 68-79) that resides in the central portion of the domain.

In contrast to the above, serovars in the C serogroup and serovars K and L3 show greater nucleotide variation in VDI than the B serogroup. Compared to C VDI, the other C serogroup serovars have 1 to 9 nucleotide substitutions in VDI that result in 0 to 4 amino acid changes. As above, the majority of these substitutions occur in the central region of the domain.

VDIs of serovars F and G are identical in their nucleotide and amino acid sequences.

2. Variable Domain II. VDII contains 63 nucleotides encoding 21 amino acids in the MOMP of serovars D, E, L1, F and G. The VDIIs of the remaining 10 serovars have 66 nucleotides encoding 22 amino acids. VDII shows more sequence variation than VDI in all serovars. As compared to serovar B, other B serogroup serovars have 2 to 19 nucleotide substitutions resulting in 2 to 8 amino acid changes. The least amount of variation occurs between serovars B and Ba in which there are two nucleotide substitutions that each result in an amino acid substitution. Most of the amino acid substitutions occur within an 8 amino acid region in the central part of the domain (residues 144 to 151).

The VDIIs of C serogroup serovars show considerable sequence variation having 4 to 8 nucleotide substitutions which result in 3 to 6 amino acid changes. A similar amount of variation was found in VDII sequences between the F and G serovars of the intermediate serogroup. Compared to serovar F VDII sequences, serovar G VDII has 6 nucleotide substitutions resulting in 3 amino acid changes.

3. <u>Variable domain III</u>. VDIII is the
smallest and least variable of the four domains.
VDIII contains 42 nucleotides which encode 14 amino
acids. The VDIII sequences of A, H, J and L3 are
identical to each other. Serovars D, L1 and F have
identical sequences which differ from serovar B by 3
amino acid substitutions. The remaining serovars
have sequences differing from serovar B by 1 to 4
amino acid substitutions.

4. <u>Variable Domain IV</u>. VDIV is the largest
variable domain containing 96 nucleotides encoding
32 amino acids in serovars A, I and all B serogroup
serovars, and 99 nucleotides encoding 33 amino acids
in the remaining serovars. The VDIV domain can be
separated into three distinct regions based on
sequence homology: (i) the N-terminal region
(residues 288-295), (ii) the central region
(residues 296-306), and (iii) the C-terminal regions
(residues 307-317).

Using micro-immunofluorescence polyclonal
mouse sera, the 15 serovars have been separated into
three serogroups. Group 1 or the B serogroup
contains serovars B, Ba, D, E, L1 and L2; group 2,
the intermediate serogroup, serovars F and G; and
group 3, the C serogroup serovars A, C, H, I, J, K
and L3.

Within the B serogroup, serovars B and Ba
have identical sequences in the N-terminal regions,
as do serovars D and E. There are three amino acid
substitutions between serovars B-Ba and D-E.
Serovars L1 and L2 contain one and two additional
amino acid substitutions in this region,
respectively. Within the C serogroup, serovars C,

H, I, J and K have identical sequences in the N-
terminal region of VDIV. Serovar L3 contains a
single amino acid substitution. Serovar A contains
the most variable sequence in this region,
5   containing 4 sequence substitutions as compared to
the prototype C serovar sequence. Serovars F and G
differ in this region by two amino acids.

The central region of VDIV is the most highly
conserved sequence among the VDs of the chlamydial
10   serovars. A nine amino acid sequence (TTLNPTIAG) is
conserved in all serovars except serovar K, which
contains a threonine (T) instead of an alanine (A)
(residue 303). The central NPT sequence in VDIV is
present in all 15 serovars.

15   The C-terminal region of VDIV is highly
conserved among the B serogroup servars. In
contrast, this region of VDIV in C serogroup
serovars shows a significant amount of amino acid
variation, although the amino acid composition is
20   similar (predominately S, T, E and A). Serovars F
and G differ by three amino acid residues in this
region.

In contrast to the historical classification
noted above, wherein serovars K and L3 are placed
25   within the C serogroup, immunotyping with monoclonal
antibodies specific to MOMP has been recently
employed and serovars K and L3 have been
reclassified to the intermediate serogroup. Based
on the present invention and comparative MOMP VD
30   amino acid homology analysis, however, serovars K
and L3 clearly belong in the historical C serogroup.

c. <u>Inter-and intra-serogroup amino acid</u>
<u>homologies.</u>

The percent of amino acid homology and the
total number of amino acid substitutions for the VDs
5   of each serovar are shown in Table 2.

TABLE 2. Comparative amino acid homologies of MOMP VDs of the 15 C. trachomatis serovars[a]

	B	Ba	D	E	L1	L2	F	G	C	A	H	I	J	K	L3
B	100[b] (0)[c]														
Ba	94.2 (5)	100 (0)													
D	80.2 (17)	77.9 (19)	100 (0)												
E	76.7 (20)	76.7 (20)	82.4 (15)	100 (0)											
L1	77.9 (19)	76.7 (20)	85.9 (12)	82.4 (15)	100 (0)										
L2	77.9 (19)	76.7 (20)	76.7 (20)	76 (21)	77.3 (20)	100 (0)									
F	43.7 (49)	43.7 (49)	44.8 (48)	39.1 (53)	44.8 (45)	40.2 (52)	100 (0)								
G	42.5 (50)	43.7 (49)	43.7 (49)	43.7 (49)	44.8 (48)	42.2 (50)	85.1 (13)	100 (0)							
C	38.2 (55)	37.1 (56)	38.2 (55)	40.4 (53)	40.4 (53)	36 (57)	34.8 (58)	39.3 (54)	100 (0)						
A	40.9 (52)	39.8 (53)	42.7 (51)	43.2 (50)	43.2 (50)	39.8 (53)	37.5 (55)	40.9 (52)	77.5 (20)	100 (0)					
H	40.4 (53)	39.3 (54)	39.3 (54)	40.2 (53)	42.7 (51)	39.3 (54)	34 (57)	38.2 (55)	83.1 (15)	74.2 (23)	100 (0)				
I	43.2 (50)	42 (51)	42 (51)	42 (51)	44.3 (49)	42 (51)	34.1 (58)	37.3 (55)	86.4 (12)	81.8 (16)	83 (15)	100 (0)			
J	42.7 (51)	41.6 (52)	41.6 (52)	43.8 (50)	46.1 (48)	40.4 (53)	37.1 (56)	40.4 (53)	84.8 (12)	84.3 (16)	86.5 (12)	87.6 (11)	100 (0)		
K	42.7 (51)	41.6 (52)	40.4 (53)	41.6 (52)	41.6 (52)	38.2 (55)	32.5 (60)	36 (57)	85.6 (10)	79.3 (22)	85.1 (15)	81.8 (16)	86.5 (12)	100 (0)	
L3	41.6 (52)	40.4 (55)	41.6 (52)	43.8 (53)	42.7 (51)	43.8 (55)	39.3 (54)	43.8 (50)	85.6 (13)	79.5 (15)	85.1 (15)	86.4 (12)	87.6 (11)	80.9 (17)	100 (0)

[a] Amino acid sequence of the four VDs were used for calculating homologies.
[b] Percent homology.
[c] Number of amino acid substitutions.

Inter-group (1, 2 and 3) VD amino acid homology is between 33 and 46%. Intra-group VD amino acid homology was between 74 and 94% for groups 1 and 3, and 85% for group 2. Therefore, classification of C. trachomatis isolates based on MOMP VD amino acid homology correlates extremely well with the historical serological classification and generally with immunotyping using MOMP specific monoclonal antibodies (see e.g., Wang et al., supra). The present inventors note, however, that recent immunotyping using monoclonal antibodies places serovars K and L3 within the intermediate serogroup (serovars F and G), whereas VD amino acid homology comparison shows serovars K and L3 to be more closely related to serovars within the C serogroup.

III. Immunogenicity based on hydrophilicity and charge values.

The VDs are hydrophilic and charged; properties indicating that these sequences are associated with the chlamydial cell surface and are potential antigenic sites. In fact, major antigenic sites which elicit the formation of protective serotyping antibodies have been mapped to VDI, VDII, and VDIV of the MOMP of serovars A, B, C and L2 (Baehr et al., supra). However, the antigenic properties of the VDs of the remaining 11 serovars have not as yet been described.

In order to define the antigenicity of these domains, the MOMP VDs of the chlamydial serovars were analyzed for hydrophilicity and charge using standard measurements and calculations. The number of charged amino acids and the hydrophilicity value of sequences within each MOMP VD for the 15

serovars was determined.   These values are provided
in Table 3.

TABLE 2.

TABLE 3. The hydrophilicity values and charges of *C. trachomatis* MOMP VDs

	VDI Basic[a]	VDI Acidic[b]	VDI HV[c]	VDII Basic	VDII Acidic	VDII HV	VDIII Basic	VDIII Acidic	VDIII HV	N' end VDIV Basic	N' end VDIV Acidic	N' end VDIV HV	Middle Basic	Middle Acidic	Middle HV	C' end of VDIV Basic	C' end of VDIV Acidic	C' end of VDIV HV
B	2	1	+1.8	2	1	+1.8	1	3	+4.3	1	2	+2.6	0	0	-5.1	1	3	+7.6
Ba	2	1	+2.8	2	1	+2.2	1	3	+5.1	1	2	+2.6	0	0	-5.1	1	3	+7.6
D	2	2	+8.8	4	1	+9.7	1	2	+3.6	1	1	+0.2	0	0	-5.1	1	3	+7.3
E	2	2	+6.4	3	0	+7.8	1	2	-3.8	1	—	+0.2	0	0	-5.1	1	3	+7.5
L1	2	1	+3.5	4	1	+12.5	1	3	+3.6	1	—	-1.9	0	0	-5.1	1	3	+7.9
L2	2	1	+3.5	4	1	+7.7	1	3	+3.7	1	1	-0.5	0	0	-5.1	1	3	+7.5
AVG[d]	2	1.3	+4.47	3.2	0.8	+6.95	1	2.8	+2.75	1	1.3	+0.53	0	0	-5.1	1	3	+7.57
F	2	2	+3.5	2	1	+2.1	1	3	+3.6	1	2	-2.5	0	0	-5.1	0	2	+1.3
G	2	2	+3.5	2	0	0	2	2	-2.8	2	1	+1.9	0	0	-5.1	0	2	+0.6
AVG[e]	2	2	+3.5	2	0.5	+1.05	1.5	2.5	+0.4	1.5	1.5	-0.3	0	0	-5.1	0	2	+0.95
C	2	1	+1.5	1	0	-0.3	2	3	-2.7	1	2	+1.1	0	0	-5.1	1	3	+6.6
A	3	1	+5.9	1	0	-3.1	2	3	+0.1	2	1	+3.0	0	0	-5.1	1	3	+6.6
H	3	4	+8.8	4	1	+5.7	1	3	+0.1	1	2	+1.1	0	0	-5.1	1	3	+6.6
I	3	2	+7.0	2	0	-0.1	1	3	-1.2	1	2	+1.1	0	0	-5.1	1	3	+6.6
J	2	1	+1.5	1	0	-7.4	1	3	+0.1	1	2	+1.1	0	0	-5.1	1	3	+6.6
K	3	2	+6.5	2	2	+0.1	1	3	-0.9	1	2	+1.1	0	0	-5.0	1	3	+7.3
L3	3	2	+6.2	2	0	-3.6	1	3	+0.1	1	2	+1.4	0	0	-5.1	1	3	+7.3
AVG[f]	2.7	1.9	+5.3	2	0.7	-1.26	1.1	2.9	-0.63	1.1	1.9	+1.4	0	0	-5.1	1	3	+6.8

[a] Total number of histidines, arginines and lysines.
[b] Total number of aspartic acids and glutamic acids.
[c] HV = hydrophilicity value assigned by Hopp and Woods, Proc. Natl. Acad. Sci. USA 78:3824-2828 (1981)
[d] Average charge and HV for the B serogroup.
[e] Average charge and HV for the intermediate serogroup.
[f] Average charge and HV for the C serogroup.

5   As the data demonstrates, the VDs have a broad range of
hydrophilicity and charge. The VDI sequences of serovars in the B
serogroup had an average hydrophilicity value of 4.47, and contained
an average of 2 positive (basic) charges and 1.3 negative (acidic)
charges, while the VDII sequences had an average hydrophilicity value
10of 6.95, and contained an average of 1.35 positive charges and 3.2
negative charges. Based on these results, the VDII sequences of
serovars within the B serogroup are predicated to have the greater
immunogenic potential between the two.

Indeed, as reported by Stephens et al. (J. Exp. Med.,
15supra) epitope mapping studies using antisera prepared against
synthetic peptides corresponding to MOMP VDs show that serovar-
specific determinants are located in the VDII of the MOMP of serovars
B, C and L2, and a highly conserved species-specific MOMP determinant
is mapped to VDIV as depicted in Table 4.

- 25b-

TABLE 4. The charge and hydrophilicity value of epitopes on *C. trachomatis* MOMP mapped by MAbs

Monoclonal antibody	Sequence of epitope[a]	Amino Acids		Hydrophilicity value	Location of epitope
		Basic	Acidic		
A-20	D-V-A-G-L-E-K-D-P-V	1	3	+4.1	A-VDI
L21-45	D-N-E-N-H-A-T-V-S-D-S-K-L-V	2	3	+6.8	L2-VDII
B-B6	N-N-E-N-Q-T-K-V-G-N-G-A-F-V	1	1	+0.9	B-VDII
B-B5	P-T-I-A-G-A-G-D-V-K-T-S-A-E-G	1	2	+3.8	B-VDIV C-terminus
L21-10	T-T-L-N-P-T-I-A-G	0	0	-5.1	VDIV of all serovars except K

[a]Sequence of epitope as determined by lambda gt11 epitope mapping according to Baehr et al., supra.

In contrast to the above, VDI sequences of C serogroup
serovars MOMP were more hydrophilic and charged than VDII.
The average hydrophilicity for VDI was 5.3 with 1.4 positive
and 2.7 negative charges, while the hydrophilicity value for

5   VDII was -1.24 with 2 positive and 0.7 negative charges.
These data suggest that VDI sequences of C serovars have a
greater immunogenic potential than VDII sequences. This
conclusion is supported by the work of Baehr et al., supra,
who, using a lambda-gt11 epitope mapping procedure showed that

10   the serovar-specific MOMP epitope for serovar A and L2 mapped
to VDI and VDII, respectively, whereas subspecies, serogroup
and species-specific epitopes mapped to VDIV. Specifically,
the epitope recognized by monoclonal antibody A-20, which is
specific for serovar A MOMP has been mapped to VDI (Table 4).

15   As with the C serogroup, the MOMP VDI sequences of the
intermediate serovars F and G were more hydrophilic and
charged than the VDII sequences. Several serovars (D, E and
H) contain both VDI and VDII sequences which are extremely
hydrophilic and charged, suggesting that in MOMPs of these

20   serovars, both domains may be immunogenic.
Although antigenic sites have not yet been mapped to
VDIII, its hydrophilicity and charge values implicate this
domain as a potential immunogenic site as well.
As noted previously, VDIV has been divided into three

25   regions, (i) N-terminal (residues 284-295), central (residues
296-306) and C-terminal (residues 307-317). The

hydrophilicity and charge for each region has been independently calculated.

The VDIV domain shows a consistent pattern of hydrophilicity and charge in each region for MOMPs of all serovars. The N-terminal region is charged but is weakly hydrophilic; the central region is consistently uncharged and is hydrophobic; and the C-terminal region is both highly charged and hydrophilic, with the exception of serovars F and G which have only moderate charge and hydrophilicity.

The single exception in correlating hydrophilicity and charge with known MOMP antigenic determinants is the species-specific MOMP epitope recognized by monoclonal antibody L21-10. This epitope is located in the central region of VDIV (TTLNPTIAG) and is uncharged and strongly hydrophobic (-5.1). From the epitope data, it appears that serovar-specific determinants are located in MOMP VDI, VDII or both, while common or serogroup-specific MOMP antigenic determinants are located in VDIV.

ABSTRACT OF INVENTION

The nucleotide and deduced amino acid sequences of the four
variable domains of the major outer membrane proteins of the
15 serovars of <u>Chlamydia</u> <u>trachomatis</u> are disclosed together
5   with sequence and immunogenic analysis of these domains.
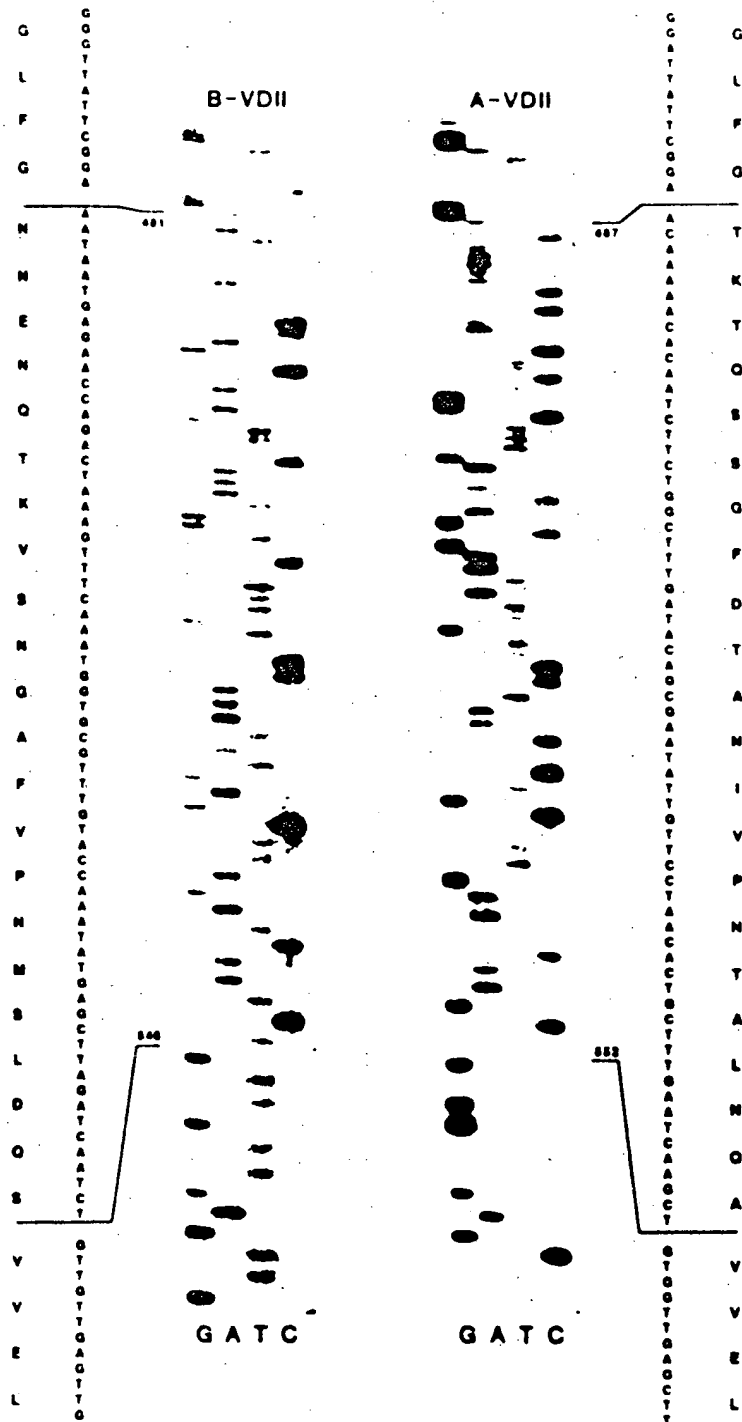
Figure 1

Location of MuMP Gene Sequences to which Synthetic Complementary Oligonucleotide were Constructed

Figure 2

Figure 3

Figure 4

B

```
55      64                    83      92                    B    132      139                      160    166
DVNKEFQMG |-AKPTTTTGNAVAPSTLT-ARE| NPAYGRHMQ       B    NLVGLFG |NNENQTKVSNGAFVPNMSLDGS| VVELYT
......... |-.....A.....T......-...| .........       Ba           |.............ST.......| ......
......... |-.....D...SA......-... | .........       D    ....... |D....KT.KAES-.....F...| ......
......... |-D..S.....T..T...-...  | .........       E    ....... |D....ST.KTNS-.........| ......
......... |-.....A.....A...C.-... | .........       L1   ....... |D....ST.KKD.-.....F...| ......
......... |-.....A....A...C.-...  | .........       L2   ....... |D...HAT..DSKL........| ......

DVNKEFEMG |EALAGASGNTTSTLSKL-VERT| NPAYGKHMQ        F    NLVGLFG |DGVNATKP-AADSIPNVQLNQS| VVELYT
                                                    G    ....... |..E...Q.-..T.........| ......

DVNKEFQMG |AAPTTSDVAGLQNEPTTNVARP| NPAYGKHMQ        C    NLVGLFG |TKTQSSSFNTAKLIPNTALNEA| VVELYT
......... |.............EK..VA....| .........        A    ....... |........G.D..NIV......Q.| .....T
......... |.......N.A.D....K.....| .........         H    ....... |...K..D......V..I...R.| .....T
......... |.......K....E.........| .........         I    ....... |.........N......V..A...Q.| .....T
......... |.....................| .........          J    ....... |........A.....N.F......Q.| .....T
......... |...........E.........| .........          K    ....... |........Y.K...N.V....DR.| .....T
......... |..E.....T...S........| .........          L3   ....... |........TN......V.....G.| .....T
```

D

```
220 224        237 241                               284 288                              317  321
GYVG |KEFPLDLTAGTDAA| TGTK           B    AQPK |SAETIFDVTTLNPTIAGAGDVKTSA-EGLP| GTMQ
.... |.........S.....| ....           Ea   .... |..TA...T..........G.-.........| ....
.... |..F..........| ....             D    .... |..TA...T..........G.-.........| ....
.... |Q.F..A.I.....| ....             E    .... |..TA...T..........A..-........| ....
.... |..F..........| ....             L1   .... |L.TA...T........E..AN.-.......| ....
.... |Q.F....K....GV| ....            L2   .... |..T.V..............A..-.......| ....

GYVG |KEFPLDLTAGTDAA| TGTK           F    AQPR |LVTPVVDITTLNPTIAGCGSVAGANTEGLS| GTMQ
.... |Q....A........| ....            G    ...K |.AK................VA..S......| ....

GYVG |AEFPLNITAGTEAA| TGTK           C    AQPK |LAEAILDVTTLNPTIAGKGSVVSACTDNELA| GTMQ
.... |.....D........| ....            A    .... |..KPV.T.........T...SA-E.....| ....
.... |.....D........| ....            H    .... |................T..AS.S..D.| ....
.... |.....D.I......| ....            I    .... |................T...SA-E....| ....
.... |.....D........| ....            J    .... |................T..AS.SE.D.| ....
.... |V....D........| ....            K    .... |.............T..A...S.S.....| ....
.... |.....D........| ....            L3   .... |....V.................AS.SE....| ....
```

Fig. 5. C. trachomatis serovar B major outer membrane protein variable domain sequences*

B-VD

256**
                                                                315
GGCCAGCCTACAACTACTACAGGCAATGCTGTAGCTCCATCCACTCTTACAGCAAGAGAG
 A K P T T T G N A V A P S T L T A R E

B-VDII

481   546
AATAATGAGAACCAGACTAAAGTTCAAATGGTGCGTTTGTACCAAATATGAGCTTAGATCAATCT
 N N E N Q T K V S N G A F V P N M S L D Q S

B-VDIII

736   777
AAGGAGTTGCCTCTTGATCTTACAGCAGGAACAGATGCTGCG
 K E L P L D L T A G T D A A

B-VDIV
928   1017
TCAGCCGAGACTATCTTTGATGTTACCACTCTGAACCCAACTATTGCTGGAGCTGGCGATGTGAAACTAGCGGCAGAGGGTCAGCTCGGA
 S A E T I F D V T T L N P T I A G A G D V K T S A E G Q L G

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amin acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 6. C. trachomatis serovar Ba major outer membrane protein variable domain sequences*

```
           256**                                           315
Ba-VDI     GCCAAGCCTACAGCTACTACAGGCAATGCTACAGCTCCATCCACTCTTACAGCAAGAGAG
            A  K  P  T  A  T  T  G  N  A  T  A  P  S  T  L  T  A  R  E

           481                                                    546
Ba-VDII    AATAATGAGAACCAGACTAAAGTTCAAATAGTACGTTTGTACCAAATATGAGCTTAGATCAATCT
            N  N  E  N  Q  T  K  V  S  N  S  T  F  V  P  N  M  S  L  D  Q  S

Ba-VDIII   736                                      777
           AAGGAGTTGCCTCTTGATCTTACATCAGGAACAGATGCTGCG
            K  E  L  P  L  D  L  T  S  G  T  D  A  A

Ba-VDIV                                                                  1017
           928
           TCAGCCGAGACTATCTTTGATGTTACCACTCTGAACCCAACTATTGCTGGAGCTGGCGATGTGAAAACTAGCGCCGAGAGGGTCAGCTCGGA
            S  A  E  T  I  F  D  V  T  T  L  N  P  T  I  A  G  A  G  D  V  K  T  S  A  E  G  G  L  G
```

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 7. C. trachomatis serovar D major outer membrane protein variable domain sequences*

D-VDI
256**
```
                                                         315
GCCAAGCCTACAACTGATACAGGCAATAGTGCAGCTCCATCCACTCTTACAGCCAAGAGAG
 A  K  P  T  T  D  T  G  N  S  A  A  P  S  T  L  T  A  R  E
```

D-VDII
481
```
                                                         543
GATAATGAAAATCAAAAACGGTCAAAGCGGAGTCTGTACCAAATATGAGCTTTGATCAATCT
 D  N  E  N  Q  K  T  V  K  A  E  S  V  P  N  M  S  F  D  Q  S
```

D-VDIII
733
```
                                          744
AAGGAGTTCCCTCTGATCTTACAGCAGGAACAGATGCTGCG
 K  E  F  P  L  I  L  Q  Q  E  Q  M  L  R
```

D-VDIV
1014
925
```
TCAGCTACAGCTATTTTGATACTACCACCGCTTAACCCAACTATTGCTGGAGCTGGCGATGTGAAACTCGCCGCAGAGGGTCAGCTCCGA
 S  A  T  A  I  P  D  T  T  T  L  N  P  T  I  A  G  A  G  D  V  K  T  G  A  E  G  Q  L  G
```

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 8. C. trachomatis serovar E major outer membrane protein variable domain sequences*

256**   315
E-VDI GACAAGCCTACAAGTACTACAGGCAATGCTACAGCTCCAACCACTCTTACAGCAAGAGAG
         D K P T S T T G N A T A P T T L T A R E

481   543
E-VDII GATAATGAAAATCAAAGCACGGTCAAACGAATTCTGTACCAAATATGAGCTTAGATCAATCT
         D N E N Q S T V K T N S V P N M S L D Q S

733   744
E-VDIII   CAAGAATTCCCTCTTGCACTCATAGCAGGAACTGATGCAGCG
            Q E F P P L A L I A G T D A A

E-VDIV   1014
925
       TCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGCTGGAGCTGGCGATGTGAAGCTACCAGCAGAGGGTCAGCTCGGA
         S A T A I F D T T T L N P T I A G A G D V K A S A E G Q L G

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 9. C. trachomatis serovar L1 major outer membrane protein variable domain sequences*

L1-VDI

256**   315
GCCAGCCTACAGCTACTACAGGCAATGCTGCAGCTCCATCCACTTGTACAGCAAGAGAG
 A K P T A T T G N A A A P S T C T A R E

L1-VDII

481   543
GATAATGAAAATCAAAGCACGGTCAAAAAGGATGCTGTACCAAATATGAGCTTTGATCAATCT
 D N E N Q S T V K K D A V P N M S F D Q S

L1-VDIII

733   774
AAGGAGTTCCCTCTTGATCTTACAGCAGGAACAGATGCTGCG
 K E F P L D L T A G T D A A

L1-VDIV
925   1014
TTAGCTACAGCTATCTTTGATACTACCACCACGCTTAACCCAACTATTGCTGGAGCTGGCGAGTGAAAGCTAACGCAGAGGTCAGCTCGGA
 L A T A I F D T T T L N P T I A G A G E V K A N A E G Q L G

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 10. C. trachomatis serovar L2 major outer membrane protein variable domain sequences*

L2-VDI   256**   315
          GCCAAGCCTACAACTGCTACAGGCAATGCTGCAGCTCCATCCACTTGTACAGCAAGAGAG
          A K P T T A T G N A A A P S T C T A R E

L2-VDII   481   546
          GATAATGAGAACCATGCTACAGTTTCAGATAGTAAGCTTGTACCAAATATGAGCTTAGATCAATCT
          D N E N H A T V S D S K L V P N M S L D Q S

L2-VDIII 736   777
          CAAGAATTCCCTCTTGATCTTAAAGCAGGAACAGATGGTGTG
          Q E F P L D L K A G T D G V

L2-VDIV
928   1017
TCAGCTACAACTGTCTTTGATGTTACCACTCTGAACCCAACTATTGCTGGAGCTGGCGATGTGAAAGCTAGCGCAGAGGGTCAGCTCGGA
S A T T V F D V T T L N P T I A G A G D V K A S A E G Q L G

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 11. _C. trachomatis_ serovar F major outer membrane protein variable domain sequences*

F-VDI

```
        256                                              318
        GAGGCTTTAGCCGGAGCTTCTGGGAATACGACCTCTACTCTTTCAAAATTGGTAGAACGAACG
          E  A  L  A  G  A  S  G  N  T  T  S  T  L  S  K  L  V  E  R  T
```

F-VDII

```
        484                                              546
        GATGGTGTAAACGCCACGAAACCTGCTGCAGATAGTATTCCTAACGTGCAGTTAAATCAGTCT
          D  G  V  N  A  T  K  P  A  A  D  S  I  P  N  V  Q  L  N  Q  S
```

F-VDIII

```
                                    777
        AAGGAGTTCCTCCTTGATCTTACAGCAGGAACAGATGCAGCG
          K  E  P  P  L  D  L  T  A  G  T  D  A  A
```

F-VDIV
928
1020

```
TTGGTAACACCTGTTGTAGATATTACAACCCTTAACCAACTATTGCAGGATGGGCAGTGTAGCTGGAGCTAACACCGGAAGGACACAGATATCT
  L  V  T  P  V  V  D  I  T  T  L  N  P  T  I  A  G  C  G  S  V  A  G  A  N  T  E  G  Q  I  S
```

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 12. _C. trachomatis_ serovar G major outer membrane protein variable domain sequences*

```
                                                                    318
G-VDI    256**
         GAGGCTTTAGCCGGAGCTTCTGGGAATACGACCTCTACTCTTCAAAATTGGTAGAACGAACG
          E  A  L  A  G  A  S  G  N  T  T  S  T  L  S  K  L  V  E  R  T
```

```
                                                             546
G-VDII   484
         GATGGTGAAAACGCCACGCAGCCTGCTGCAACAAGTATTCCTAACGTGCAGTTAAATCAGTCT
          D  G  E  N  A  T  Q  P  A  A  T  S  I  P  N  V  Q  L  N  Q  S
```

```
                                              777
G-VDIII  736
         CAAGAATTCCCTCCTTGCACTCACAGCAGGAACTGATGCAGCG
          Q  E  F  P  P  L  A  L  T  A  G  T  L  A  A
```

```
G-VDIV                                                                        1020
928
TTGGCAAAACCTGTTGTAGATATTACAACCCTTAACCCAACTATTGCAGGATGCGGCAGTGTAGCTGCAGCTAACTCCGAAGGACAGATATCT
 L  A  K  P  V  V  D  I  T  T  L  N  P  T  I  A  G  C  G  S  V  V  A  A  N  S  E  G  Q  I  S
```

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 13. C. trachomatis serovar C major outer membrane protein variable domain sequences*

C-VDI
```
      256**                                                       321
      GCGGCGGCCTACTACCAGCGATGTAGCAGGCTTACAAAACGATCCAACAACAAACGTTGCTCGTCCA
      A  A  P  T  T  S  D  V  A  G  L  Q  N  D  P  T  T  N  V  A  R  P
```

C-VDII
```
      487                                                              552
      ACAAAACACAATCTTCTAGCTTTAATACAGCGAATATTATTCCTAACACTGCTTGAATGAAGCT
      T  K  T  Q  S  S  S  F  N  T  A  K  L  I  P  N  T  A  L  N  E  A
```

C-VDIII
```
      742                                            763
      GCGGAATTCCACTAATATTACCGCAGGAACAGAAGCTGCG
      A  E  F  P  L  N  I  T  A  G  T  E  A  A
```

C-VDIV
```
934                                                                        1026
TTGGCTGAAGCAATCTTGGATGTCACTACTCTGAACCCGACTATCGCTGGTGGTAAAGGAAGTGTGTCTCTGCCGAACCGATAACGAACTGGCT
L  A  E  A  I  L  D  V  T  T  L  N  P  T  I  A  G  K  G  S  V  V  S  A  G  T  D  N  E  L  A
```

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 14. C. trachomatis serovar A major outer membrane protein variable domain sequences*

A-VDI
```
                                                                              321
     256*
     GCGGCGCCTACTACCAGCGATGTAGCAGGCTTAGAAAACGATCCAGTAGCAAATGTTGCTCGCCCA
      A  A  P  T  T  S  D  V  A  G  L  E  K  D  P  V  A  N  V  A  R  P
```

A-VDII
```
                                                             552
     487
     ACAAAAACACAATCTTCTGGCTTGATACAGCGAATATTGTTCCTAACACTGCTTGAATCAAGCT
      T  K  T  Q  S  S  G  F  D  T  A  N  I  V  P  N  T  A  L  N  Q  A
```

A-VDIII
```
                                      783
     742
     GCGGAATTCCACTTGATATTACCCCAGAACAGAAGCTGCG
      A  E  F  P  L  D  I  T  A  G  T  E  A  A
```

A-VDIV
```
                                                                                                1023
934
TTGGCTAAACCAGTCTTGGATACCACTACTCTAAACCCGACCATCGCTGGTAAAGGAACTGTGGTCTCTTCCGCAGAAAACGAACTGGCT
 L  A  K  P  V  L  D  T  T  T  L  N  P  T  I  A  G  K  G  T  V  V  S  S  A  E  N  E  L  A
```

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 15. _C. trachomatis_ serovar H major outer membrane protein variable domain sequences*

321

H-VDI
256**
GCGGCGCCTACTACCAACGATGCAGCAGACTTACAAAACGATCCAAAACAAATGTTGCTCGTCCA
A A P T T N D A A D L Q N D P K T N V A R P

552

H-VDII
487
ACAAAACAAATCTTCTGATTTAATACAGCGAAGCTGTTCCTAACATTGCTTGAATCGAGCT
T K T K S S D F N T A K L V P N I A L N R A

783

H-VDIII
742
GCGGAATTCCACTTGATATTACCGCAGGAACAGAAGCTGCG
A E F P L D I T A G T E A A

1026

H-VDIV
934
TTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCGCTGGTAAAGGAACTGTGGCCGCTTCCGGAAGCGATAACGACCTGGCT
L A E A I L D V T T L N P T I A G K G T V V A S G S D N D L A

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 16. *C. trachomatis* serovar I major outer membrane protein variable domain sequences*

```
I-VDI
           256**                                                              321
           GCGGCGCCTACTACCAAGGATGTAGCAGGCTTAGAAAACGATCCAACACACAAATGTTGCTCGTCCA
            A   A   P   T   T   K   D   V   A   G   L   E   N   D   P   T   T   N   V   A   R   P

I-VDII
           487                                                                552
           ACAAAACACAATCTTCTAACTTTAATACAGGGAAGCTTGTTCCTAACGGTGCTTTGAATCAAGCT
            T   K   T   Q   S   S   N   F   N   T   A   K   L   V   P   N   A   A   L   N   Q   A

I-VDIII
           742                                         783
           GCGGAATTCCACTTGATATTATGCGCAGGAACAGAAGCTGCG
            A   E   F   P   P   L   D   I   I   A   G   T   E   A   A

I-VDIV
934                                                                                        1023
TTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCGCTGGTAAGGAACTGTGTCTTCCGCAGAAAACGAACTGGCT
 L   A   E   A   I   L   D   V   T   T   L   N   P   T   I   A   G   K   G   T   V   V   S   S   A   E   N   E   L   A
```

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 17. *C. trachomatis* serovar J major outer membrane protein variable domain sequences*

-VDI

256**   321
GCGGCGGCCTACTACCAGGCGATGTAGCAGGCTTACAAACGATCCAACAACAAATGTTGCTCGTCCA
 A A P T T S D V A G L Q N D P T T N V A R P

-VDII

487   552
ACAAAACACAAGCTTCTAGCTTTAATACAGCGAATCTTTTCCTAACACTGCTTTGAATCAAGCT
 T K T Q A S S P N T A N L P P N T A L N Q A

-VDIII

742   783
GCGGAATTCCACTTGATATTACCGCAGGAACAGAAGCTGCG
 A E F P L D I T A G T E A A

-VDIV
34   1026
TTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCGCTGGTAAAGGAACTGTGGTCGCTTCCGGAAGCGAAAACGACCTGGCT
 L A E A I L D V T T L N P T I A G K G T V V A S G S E N D L A

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

*Location of nucleotide sequence within the MOMP gene.

Fig. 18. C. trachomatis serovar K major outer membrane protein variable domain sequences*

K-VDI
                                                                321
256**
GCGGCGCCTACTACCAGCGGATGTAGAAGGCTTACAAAACGATCCAACAACAAATGTTGCTCGTCCA
A A P T T S D V E G L Q N D P T T N V A R P

K-VDII
                                                                552
487
ACAAAACACAATATTCTAAGTTTAATACAGCGAATCTGTTCCTAACACTGCTTTGGATCGGAGCT
T K T Q Y S K F N T A N L V P N T A L D R A

K-VDIII
                                    783
742
GTGGAATTCCACTTGATATTACCGCAGGAACAGAAGCTGCG
V E F P L D I T A G T E A A

K-VDIV
                                                                        1026
934
TTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCACCATCACTGGTAAGGGAGCTGTGGTCTCTTCCGGAAGCGGATAACGAACTGGCT
L A E A I L D V T T L N P T I T G K G A V V S S G D N E L A

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid
sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.